Development of a genre-specific electronic dictionary and automatic gloss-embedding system
-- Converting Internet English resources to teaching materials on demand --

Hiroshi OHTAKE
Kyoto Prefectural University of Medicine
ohtake@koto.kpu-m.ac.jp

Takeshi KAWAMOTO
Hiroshima University Graduate School of Biomedical Sciences
tkawamoto@hiroshima-u.ac.jp

Masataka TAKEKOSHI
Tokai University School of Medicine
mtakekos@is.icc.u-tokai.ac.jp

Masako KUNIMURA
Fukui National College of Technology
msong@po.mitene.ac.jp

Brian MORREN
Fukui Prefectural University
morren@fpu.ac.jp

Hiroaki TAKEUCHI
Shizuoka University
sbhtake@ipc.shizuoka.ac.jp

Yoshihiro UGAWA
Miyagi University of Education
ugawa@ipc.miyakyo-u.ac.jp

Nobuyuki FUJITA
National Institute of Genetics
nfujita@lab.nig.ac.jp

Shuji KANEKO
Graduate School of Pharmaceutical Sciences
Kyoto University
skaneko@pharm.kyoto-u.ac.jp

Abstract

We have created a genre-specific electronic English-Japanese dictionary and an automatic gloss-embedding system for electronic English texts in a specific discipline (life sciences and related issues). This particular dictionary and the automatic gloss-embedding system has made it possible for English resources on the Internet to be instantaneously converted to a hyper-text type electronic file. In this way, we hope that authentic English materials that are widely available on the Internet may be readily adopted for the purpose of learning/teaching English. The provision of a gloss for a particular text is expected to greatly reduce the burden on learners arising from their vocabulary deficiencies and encourage them to read relatively difficult English materials on their own. To show the convenience and effectiveness of the dictionary and the actual system itself, we have opened a home page (http://lsd-rap.pharm.kyoto-u.ac.jp) from which we hope potential users (learners/teachers) can gain a clear idea of how the whole system works. So far, we have made available more than a hundred English articles pertaining to life-sciences published in major magazines, newspapers, and journals, together with a particular gloss for each respective article.

The system is designed to be versatile so that it can also be used by learners or teachers of English from different language backgrounds as long as they compile an electronic bilingual dictionary like the one we have created. It should be noted that this system has made it possible for teachers of English to acquire and provide reading materials in the form of hyper-text on the computer display by just a few clicks of a mouse button.

1 Introduction

In recent years, Japanese teachers of English have become more interested in using articles published by newspapers, magazines such as Time and Newsweek, as well as the BBC and CNN for the purpose of providing authentic materials to their reading class students. It has even become possible to access literary works and novels on the Internet, and these can also be regarded as a useful source of teaching materials. Although electronic English texts released on the Internet appear to be accessible and readily available, there are a number of obstacles to instantly adopting them as reading materials, which has traditionally resulted in forcing teachers to use printed materials in the form of textbooks with glosses attached. However, the articles that appear in these con-

ventional types of printed textbooks tend to be out of date and the time lag sometimes causes learners to lose interest in them. This has created a dilemma for teachers in that the articles in textbooks often lack timeliness and do not attract learners very much in spite of their authenticity, while those obtained via the Internet are up-to-date but difficult for most learners to read owing to their lack of vocabulary.

In this regard, through the development of our new gloss-embedding system that can automatically insert a gloss for a particular word or phrase, teachers are able to make immediate use of these materials as they appear on the Internet. Certainly, such a computerized system is expected to overcome some of the drawbacks of traditional dictionary look-up. For example, consulting a dictionary may sometimes be misleading in that learners may misinterpret dictionary entries and fail to select the meaning that is most appropriate in the context (Bensoussan, Sim, & Weiss, 1984). This may be due to difficulties in selecting the appropriate meaning of a particular word from among all the possible meanings listed. It would also appear that dictionary use tends to result in slower processing of text in that learners with access to a dictionary require more time to complete a task than those without a dictionary (Nesi & Meara, 1991). This may then interrupt the reading flow and have an adverse effect on learners' comprehension of text.

Furthermore, a computerized glossing system is inherently more flexible than the conventional system of using marginal glosses in that it may accommodate learners of differing levels of proficiency and learning styles. For example, perceptions of word difficulty will naturally vary from person to person where what is considered difficult by some will not be considered problematic by others. This may therefore result in the glossing of already familiar words and expressions while more difficult items may not receive a gloss. Consequently, where such glosses give information that is already known to learners, this may simply serve as a distraction and thereby impede their processing. Where certain information is required by learners in order to make sense of a text, the omission of a gloss may also lead to textual misinterpretation (Davis, 1989). However, in the case of the computerized version, the glosses remain invisible unless activated by the learner so that if a word is already familiar, their attention will not be drawn to a gloss (unlike glosses printed in the margin) and their processing will remain uninterrupted. It therefore depends entirely on the learner whether a particular gloss is accessed or not, which accords well with the recent trend toward greater learner autonomy.

Overall, the computerized gloss-embedding system may also help to make the reading process less of an ordeal in that the glosses may be accessed by the click of a mouse button, thereby avoiding the time-consuming nature of dictionary look-up. It is also possible to offer a far greater number of glosses than would be feasible with the conventional type of marginal glossing, thus ensuring more complete coverage of words thought likely to present problems for learners. In this regard, the superior speed and considerable information storage capacity of computers may be seen to have certain advantages over traditional dictionaries and marginal glossing techniques (Leffa, 1992).

2 Genre specific bilingual electronic dictionary and gloss-embedding system

2.1 Creating an electronic English-Japanese dictionary relating to medical and life sciences

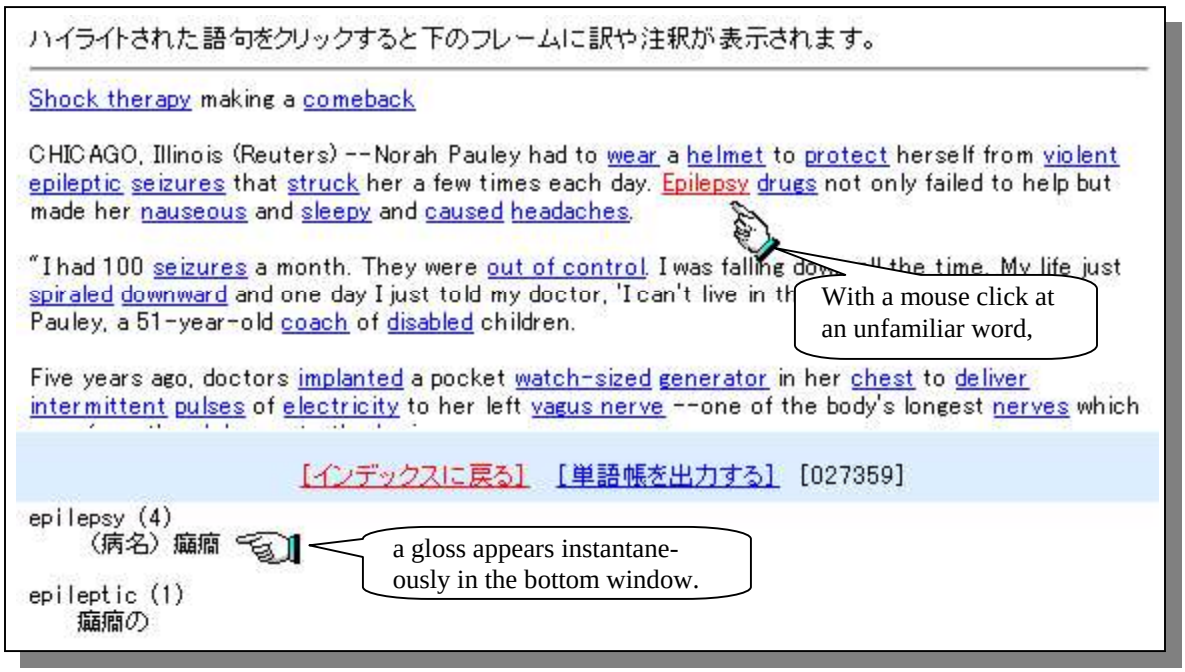
We collected a large number of articles pertaining to medical and related life sciences with the aim of creating a special electronic dictionary. The articles gathered were statistically analyzed by computer and the most frequently occurring words were selected for dictionary entries with the exclusion of basic English words which are normally learned at the threshold level. The total number amounted to approximately 8,000 words and for each word a Japanese equivalent was registered with special attention given to the particular discipline. The created dictionary was named the RAP Dictionary after our project name (Reading Assistant Project).

2.2 Development of a gloss-embedding program

A special program was developed in order to convert the original electronic English text into a hyper-text type file with glosses embedded for apparently difficult words in the text with the help of the RAP Dictionary. With this program (EtoJ Vocabulary) and the original dictionary, almost every electronic article relating to life sciences can be converted to a ready-made text for classroom use with just a few clicks of a mouse button.

Hitherto, it has been difficult for language teachers to select the particular words that are likely to cause problems for students because of their varying vocabulary levels. However, this system frees teachers from having to decide which words require a gloss and which do not, because they can be quite flexible in choosing owing to

the invisible nature of the glosses and where the decision to retrieve a gloss entirely depends on each learner. Thus, this system can convert an authentic text to reading materials in the classroom for learners with quite different levels of vocabulary.



(**Figure 1:** Sample of Converted Text with Glosses Embedded CNN030524)

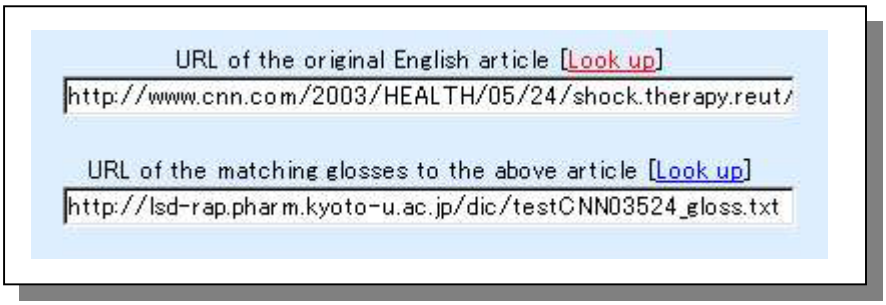
Nevertheless, although the RAP Dictionary and EtoJ Vocabulary can produce a list of particular words that are included both in a given article and the Dictionary, it does not necessarily guarantee that this system will always supply an appropriate equivalent for each English word. Therefore, teachers are expected to check the list of vocabulary produced and give an appropriate definition for some words and adjust the difficulty level of the word list accordingly. For this purpose, another program for producing a particular word list for a given article was developed, and this program also allows teachers themselves to embed the explanation of the article or give additional information.

2.3 Homepage for reading materials on demand

To show the effectiveness of this system, we have opened a homepage (http://lsd-rap.pharm.kyoto-u.ac.jp) and made the whole system available to the public. Once a week, we select one article relating to life sciences which can be obtained on the Internet and convert it into ready-to-use teaching materials for reading classes. This is done by taking advantage of the system we have developed and examining the appropriateness of the Japanese translation equivalents and replacing inappropriate definitions automatically embedded with more appropriate ones where necessary. We are hoping that a growing number of teachers will come to appreciate the convenience and effectiveness of the system. Since we opened the homepage in July 2001, more than a hundred articles have been introduced and over 50,000 visits to the site have been recorded.

2.4 The copyright issue

It is important not to violate the copyright of articles when using them even though they seem to be resources that are freely available. In view of this, we have introduced a system whereby we do not have to store a given article in our own server computer and instead invite the users themselves to directly access the original article through our own automatic conversion system on the understanding that when users make use of the articles, they do so on their own responsibility and in accordance with copyright law. To ensure that the users are fully aware that they are responsible for using a given article that may be protected under international copyright law, they are invited to press a special “consent” button prior to converting a given article into a hyper-text style text.



(**Figure 2:** Assurance of Copyright Protection -- The universal resource location of the original text is explicitly given and the user is invited to convert the text in accordance with copyright protection law.)

2.5 The versatility of the system

As illustrated here, with the dictionary and the programs we have developed, it is possible to instantly convert life science related texts into ready-made teaching materials. However, the whole system itself is sufficiently versatile that it can be applied to other reading materials belonging to other disciplines as long as users develop a particular bilingual electronic dictionary for texts from specific genres, regardless of the languages used for translation equivalents. In actual fact, we have also developed a larger dictionary (Life Science Dictionary) consisting of approximately 40,000 words, most of which are technical terms pertaining to life sciences, and a further specific dictionary for articles dealing with political, economical, and social themes. With the latter dictionary, we can use news articles appearing in major newspapers and journals as teaching materials with automatically embedded glosses. In this case, teachers should be aware that some inappropriate equivalents for English words may be included because of the automatized nature of the system. As long as teachers keep this in mind, the converted version of the original article can be appropriately used in their classrooms. This system also enables teachers to use their own preferred dictionaries and articles of their own choice in meeting their teaching goals.

2.6 System evaluation from the pedagogical perspective

The major difference between teaching materials provided in the form of conventional type printed texts and those offered through the EtoJ Vocabulary system is that learners are no longer confronted with unnecessary glosses in the new system. Learners are able to simply click a mouse button whenever they encounter a particular word or phrase that is unfamiliar to them. As previously mentioned, while level of vocabulary will vary from one student to another, teachers will no longer have to concern themselves with which particular level they should focus on since gloss consultation is left completely up to the learners themselves in this system which can embed a gloss for virtually every word. Furthermore, learners may benefit from the greater independence gained from being able to read on their own away from the ordinary classroom environment (Nation, 1990).

Through actual classroom use of the materials produced by this system, we have tried to evaluate the effectiveness of this type of innovative reading material in university reading classes. For this purpose, a questionnaire consisting of ten questions was administered to the students at the end of the course. As for two items in the questionnaire relating to “convenience” and “advantage over conventional types of printed dictionaries,” more than 80% of respondents chose 4 or 5 out of a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). This indicates that they found the material provided through this system more readable than the one in the original format. Some of the respondents explicitly remarked that without such a system they would have almost given up attempting to read challenging materials that were full of unknown words. It would appear that their affective filter was thereby lowered, enabling them to tackle seemingly difficult texts without undue anxiety (Krashen, 1982). However, a few students commented that long exposure to the reading materials displayed on the computer screen sometimes led to tiredness especially with respect to the eyes.

3 Conclusion

As far as we know, there seem to be very few non-commercial world-wide-web sites that aim to help learners improve their English reading skills, even though there are a growing number of www commercial sites advertising their services in assisting language learners. Now that computers are able to perform a similar type of media role as television or radio, the purpose of this project is to fully exploit the computer and network system in meeting the needs of the current age. In this connection, modern communication technology such as e-mail has been adopted in the field of English teaching when students are required to submit essays and reports, and the use of the Internet for educational purposes is ever expanding. In view of this, it is necessary for language teachers to seek to improve their own way of teaching and at the same time examine the effectiveness of new teaching styles in the light of language acquisition theories. There are also some pedagogical points to be clarified such as the use of electronic dictionaries and the educational effect of learners encountering a large number of unknown words in a short period of time.

In this respect, it has been found that learners are in favor of computerized glossing as it is both convenient and easy to access. Furthermore, it is generally preferred to the normal print format in that it is less distracting and causes less interruption in their reading (Davis & Lyman-Hager, 1997). Teachers also no longer have to worry about which items should be selected for glossing or what may be the optimal amount of glossing for any given text. At the same time, however, although the computerized glosses have been well-received by both learners and teachers, this does not necessarily mean that such a system should be used to the exclusion of other perhaps less popular but equally effective approaches to vocabulary learning. Computer glossing may be initially attractive in comparison with the more effortful strategy of consulting a dictionary, but an over-reliance on the computer may obscure other potentially effective ways of deriving meaning from text. Moreover, it has been found that learners tend to overestimate their degree of comprehension, perhaps because of the ease with which relevant information can be accessed, which may lead to an inflated sense of their ability to comprehend text. In this sense, a plurality of approaches is certainly likely to be of greater benefit to learners in that they may serve to complement and reinforce each other (Jacobs, Dufon, & Fong, 1994). With these important pedagogical points in mind, therefore, we would like to refine our system so that we may make a greater contribution to the modernization of English teaching in Japan.

References

Bensoussan, M., Sim, D., & Weiss, R. 1984. The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations. *Reading in a Foreign Language, 2(2),* pp. 262-276.

Davis, J. 1989. Facilitating effects of marginal glosses on foreign language reading. *The Modern Language Journal, 73(1),* pp. 41-48.

Davis, J., & Lyman-Hager, M. 1997. Computers and L2 reading: Student performance, student attitudes. *Foreign Language Annals, 30*(1), pp. 58-72.

Jacobs, G., Dufon, P., & Fong, H. 1994. L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading, 17,* pp. 19-28.

Krashen, S. 1982. *Principles and Practice in Second Language Acquisition.* Pergamon.

Leffa, V. 1992. Making foreign language texts comprehensible for beginners: An experiment with an electronic glossary. *System, 20(1),* pp. 63-73.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. Newbury House.

Nesi, H., & Meara, P. 1991. How using dictionaries affects performance in multiple-choice EFL tests. *Reading in a Foreign Language, 8(1),* pp. 631-643.